

Better Understanding of Basic Mean Square Estimation

Chieh Wu

Dedicated to Fred Daun

October 1, 2004

Abstract

The purpose of this paper is to provide a very fundamental and working understanding of Mean Square Estimation. Although very little knowledge about Mean Square Estimation is needed to understand the paper, a strong probability background is necessary. Through out the paper, it will use concept such as **mean**, **variance**, **covariance** freely.

This paper will most benefit 3 types of people.

1. If you are a student and need a quick understand because the exam is tomorrow.
2. If you are a working engineer that needs a quick understanding and move on.
3. If you are a nerd like me that write these kind of paper for fun.

The first section will first provide a non-mathematically explanation of MSE. Then it will go through the mathematical derivation step by step. The last section, goes over an actually example that estimates a series of points using MSE. The example will be done in Matlab, and the source code will be provided.

From the sound of it, Mean Square Estimation seems like something extremely complicated. But its original purpose is actually pretty plain and simple. You are essentially using what you know to make the best guess on what you don't know. For example, if we have a round object in front of us that we have never seen before, we might want to compare it to a ball. The ball in this case would be our estimation. We are using a ball because they are some kind of similarity, in this case they are both round.

Now instead of using something constant as an example, let's use something that's much more complicated, a woman. Let's call her Sarah. And let us assume that a man is making the estimation, so the best estimation object they can pick is another fellow man. We will call this man Fred. In many cases, this might be a fatal estimator but since they are both have two eyes we at least have a starting point.

When I originally chose a woman as a less constant object I meant emotionally. Not that I was trying to say that woman are emotionally unstable, but rather that they have a higher variance emotionally compare to a ball. Sometimes our Sarah is happy, and sometime she is sad. It is never constant and during the course of a day, anything could happen. However, if we were to randomly bump into Sarah, we would expect to see her relatively happy. This is the expectation of our estimatee. As for our estimator Fred, he is also not constant. He also goes through all the emotions that Sarah goes through. However since its much harder for Fred to change, his emotional variance is much lower compare to Sarah. If we were to bump into Fred during the day, we would expect to see Fred relatively happy just like Sarah. That is the expectation of Fred.

If we were to graph the pdf of both Sarah(s) and Fred(F), we might get a graph that looks like this.

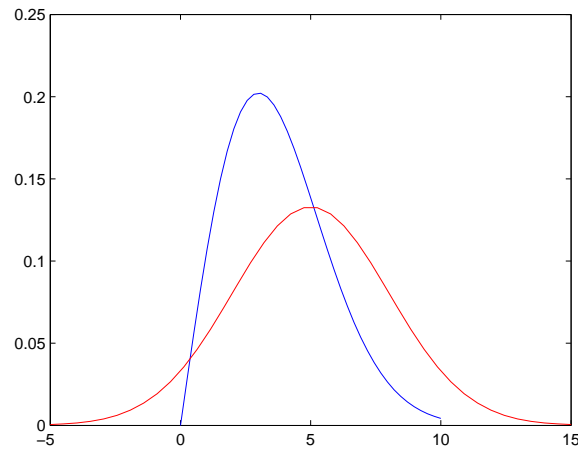


Figure 1: The emotional state of both Sarah and Fred. Sarah is denoted as the red graph while Fred is the blue graph. Notice that Sarah has a much higher variance compare to Fred. The expectation for Sarah is 5 while for Fred is about 3.75

Now if we want to use Fred(F) to estimate Sarah(s), there are two pieces of information we would need to know. First, we want to know what we can multiply to make Fred look relatively like Sarah. So that if Fred is a function of F, we want to find α such that αF will look similar to Sarah. Once you have found α , the graph would then look something like this.

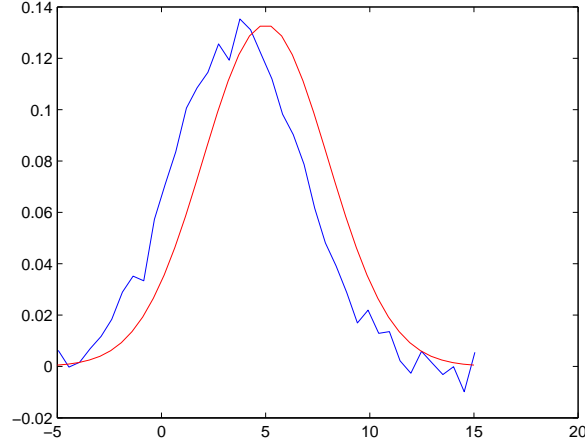


Figure 2: The emotional state of both Sarah and Fred after we have multiplied Fred. The expectation for Sarah is still 5 while for Fred is still about 3.75. Notice that the only difference is that Fred is now also a Gaussian RV with noise

This constant α will determine how we can make Fred's graph look like Sarah's. So how do we think about what this α is? To start, we would first need to know what their relationship is. But what should we use to compare? After all both Sarah and Fred change constantly. Could we use Sarah's good day to compare with Fred's bad day? The answer is to compare everything. You want to compare how Sarah change everyday to Fred's daily variance. In probability is this called the covariance.

Now that we have their relationship, if we divide this value to Fred's own personal variance we will get the alpha value. With α times Fred, we can generate the ugly estimation we see above.

$$\frac{\text{The relationship between Fred and Sarah}}{\text{Fred's variance}} = \alpha$$

This could also be written as:

$$\alpha = \frac{COV(F, s)}{Var(F)}$$

The second part is to shift the graph of Fred to the same position as Sarah. This is a much easier job to understand. We are essentially adding a constant to our previous conversion.

Before the shift

$$estimation = \alpha X$$

Now we want to add a constant to shift our estimation. We'll call it beta.

$$estimation = \alpha X + \beta$$

To find this beta is very easy. If you look at the previous graph, you would notice that the mean of Sarah is 5 while the mean for Fred is about 3.75. Their difference is 1.25. So β in this case is 1.25. But conceptually, we are taking the mean of Sarah to subtract the mean of our estimation.

$$\beta = \text{Sarah's average} - \text{Our estimator's average}$$

If we call Sarah's distribution s , this would be the equation to beta

$$\beta = E[s] - \alpha E[F]$$

The final graph would look something like this: The final equation to trans-

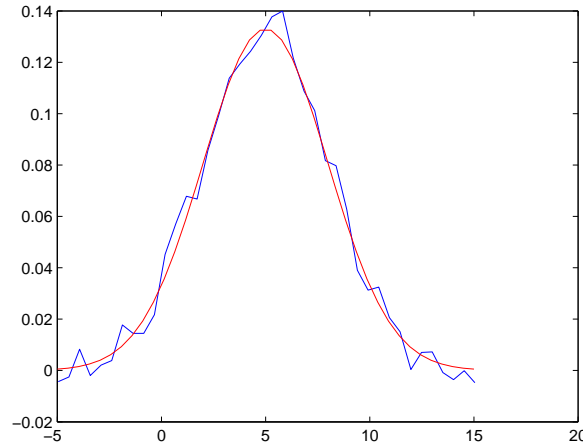


Figure 3: The emotional state of both Sarah and Fred after we have multiplied and shifted. The expectation for both Fred and Sarah is 5. Notice how Fred became a relative close estimation of Sarah

form to Sarah from Fred is:

$$s = \alpha X + \beta$$

$$s = \frac{COV(s, F)}{Var(F)} F + E[s] - \frac{COV(s, F)}{Var(F)} E[F]$$

From this equation, it means that we can use any equation to estimate. The closer the estimator the smaller the error we will get. In the next section, we will mathematically re-derive the previous conclusion and provide an example of trying to fit a function to a bunch of dots.

If you only care about the implementation of MSE, this section is really not necessary for that purpose. You may want to jump to the next section where we provide an example. This section we will derive the exact same equation we derived previously mathematically. The purpose of this section is to provide a solid mathematical background to how the previous equation came about.

Let us have a function Y as the function that we need to estimate. To estimate it, we provide another function X as a close approximation of Y . Our goal is to tweak X so that it would become closer to Y such that.

$$Y \approx \alpha X + \beta$$

In this case, we know what X is and we want to use the best α and β that would modify X to become Y . Obvious the best estimation we could use is if :

$$Y = \alpha X + \beta$$

and the error would be

$$error = Y - (\alpha X + \beta)$$

What we want to do is to choose the best α and β to minimize the error. Of course it is normally much easier to work with positive errors so we tend to use the second power of the error as error.

$$error = [Y - (\alpha X + \beta)]^2$$

Since the function Y and X are functions instead of constant values, we want to find on average what the error would be. The average is also known as the expectation of the error.

$$E[error] = E[(Y - \alpha X - \beta)^2]$$

With this expectation, we want to minimize this value. In calculus, when we want to find either the max or min of a function, we take the derivative and equate it to 0. We can use the same method here to find the minimum error. Let us first find the minimum β

$$0 = \frac{d}{d\beta} E[(Y - \alpha X - \beta)^2]$$

Since the derivative is linear, we can move it directly inside the expectation.

$$0 = E\left[\frac{d}{d\beta}(Y - \alpha X - \beta)^2\right]$$

If we were to continue with the math:

$$0 = E[2(Y - \alpha X - \beta)(-1)]$$

We can move constants outside of the expectation.

$$0 = -2E[(Y - \alpha X - \beta)]$$

The -2 will divide out with the zero and we will split up the expectations.

$$0 = E[Y] - \alpha E[X] - \beta$$

Now we can see that the best β is:

$$\beta = E[Y] - \alpha E[X]$$

Now if we substitute this back into the original equation for β we have:

$$E[error] = E[(Y - \alpha X - E[Y] + \alpha E[X])^2]$$

Now we repeat the same process, except this time we find the optimum α

$$0 = E\left[\frac{d}{d\alpha}(Y - \alpha X - E[Y] + \alpha E[X])^2\right]$$

We find the derivative:

$$0 = E[2(Y - \alpha X - E[Y] + \alpha E[X])(-X + E[X])]$$

$$0 = -2E[(Y - \alpha X - E[Y] + \alpha E[X])(X - E[X])]$$

We can rearrange the variables within the expectation and get:

$$0 = E[(Y - E[Y])(X - E[X]) - \alpha(X - E[X])^2]$$

From this we can see that the first term is the covariance while the second term is the variance.

$$0 = COV(X, Y) - \alpha Var(X)$$

Now if we rearrange the algebra to find the α we would get:

$$\alpha = \frac{COV(X, Y)}{Var(X)}$$

So finally we can estimate Y by using the equation:

$$Y \approx \frac{COV(X, Y)}{Var(X)}X - E[Y] + \frac{COV(X, Y)}{Var(X)}E[X]$$

Notice that we have derived the same equations from previously.

In this section we will go through a step by step example.

Let us have a set of dots that we want to find the best fitting equation to.

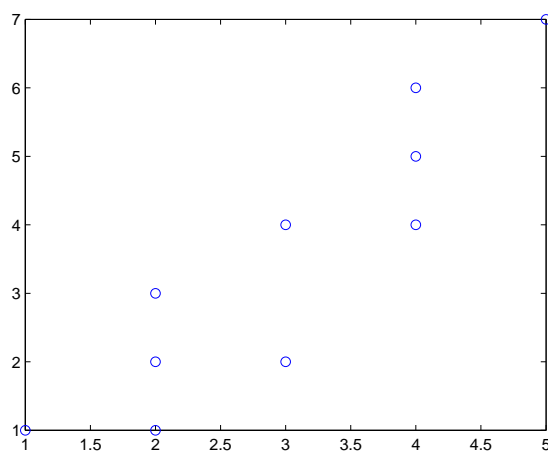


Figure 4: An example of dots that might be related with a certain function

These are the points to the set of points.

$$\begin{bmatrix} a & b \\ 1 & 1 \\ 2 & 1 \\ 2 & 2 \\ 2 & 3 \\ 3 & 2 \\ 3 & 4 \\ 4 & 4 \\ 4 & 5 \\ 4 & 6 \\ 5 & 7 \end{bmatrix}$$

We are going to use the equation x to estimate these dots

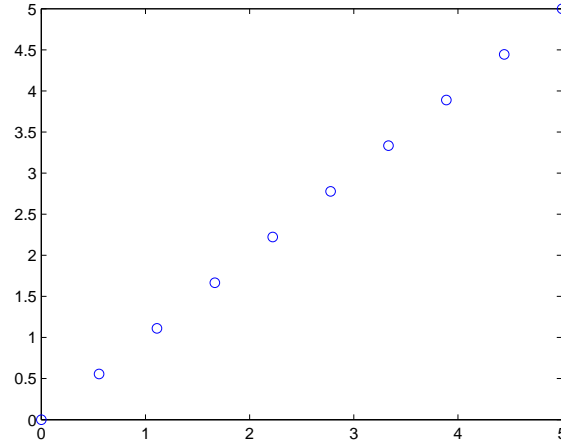


Figure 5: The estimator, these are uniformly chosen across the line x

These are the points to the estimating dots

c	d
0	0
0.5556	0.5556
1.1111	1.1111
1.6667	1.6667
2.2222	2.2222
2.7778	2.7778
3.3333	3.3333
3.8889	3.8889
4.4444	4.4444
5.0000	5.0000

The reason why we are using dots of the line instead of the actual line is because we need to have the same elements for the estimator and the estimatee to calculate the covariance. Notice that we have two columns to estimate. The x axis and the y axis. We will do them individually as parametric equations. Namely, I will use the first column of the estimator to estimate the first column of the estimatee. The same method will be used for the second column. Together they form two separate parametric equations which they would be combine in the end.

Now we can follow the equation we have derived previously to solve for the best fit line. We'll start with the first columns.

$$Y \approx \frac{COV(X,Y)}{Var(X)}X - E[Y] + \frac{COV(X,Y)}{Var(X)}E[X]$$

Before we jump into finding the equation we need to first find out many characteristics about the two columns. For the first column we have:

$$\begin{bmatrix} a \\ 1 \\ 2 \\ 2 \\ 2 \\ 3 \\ 3 \\ 4 \\ 4 \\ 4 \\ 5 \end{bmatrix}$$

expectation $E = 3$

$$a - E[a] = \text{element} - E[a] = \begin{bmatrix} a \\ 1 \\ 2 \\ 2 \\ 2 \\ 2 \\ 3 \\ 3 \\ 3 \\ 4 \\ 4 \\ 4 \\ 4 \\ 5 \end{bmatrix} - \begin{bmatrix} a \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 \\ -1 \\ -1 \\ -1 \\ -1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 2 \end{bmatrix}$$

We will now use the same method to find out the same information for the estimator.

$$\begin{bmatrix} 0 \\ 0.5556 \\ 1.1111 \\ 1.6667 \\ 2.2222 \\ 2.7778 \\ 3.3333 \\ 3.8889 \\ 4.4444 \\ 5.0000 \end{bmatrix}$$

Expectation $E = 2.5$
Variance $\text{Var} = 2.829$

$$c - E[c] = \begin{bmatrix} -2.5000 \\ -1.9444 \\ -1.3889 \\ -0.8333 \\ -0.2778 \\ 0.2778 \\ 0.8333 \\ 1.3889 \\ 1.9444 \\ 2.5000 \end{bmatrix}$$

To find the covariance for X and Y we know that:

$$COV(a, c) = E[(a - E[a])(c - E[c])]$$

Now if we want to find the covariance, it is

$$Cov(a, c) = mean \left[\begin{bmatrix} -2 \\ -1 \\ -1 \\ -1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 2 \end{bmatrix} * \begin{bmatrix} -2.5000 \\ -1.9444 \\ -1.3889 \\ -0.8333 \\ -0.2778 \\ 0.2778 \\ 0.8333 \\ 1.3889 \\ 1.9444 \\ 2.5000 \end{bmatrix} \right]$$

$$COV(a, c) = 1.833$$

Now we can solve for Y with this equation.

$$a \approx \frac{COV(a, c)}{Var(c)}c - E[a] + \frac{COV(a, c)}{Var(c)}E[c]$$

Since we are using parametric equations, lets use t as a dummy variable for the x axis. The result would be

$$a = .6480t + 1.38$$

If we were to follow the same procedure for the y axis, we would get this formula.

$$b = 1.0702t + .8245$$

If we were to combine the two parametric equations and plot out the result we will have the following graph.

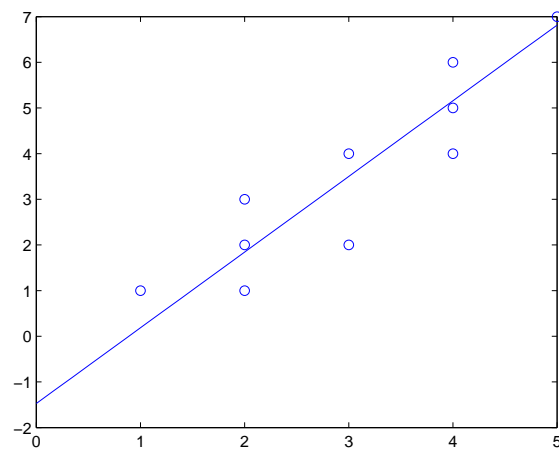


Figure 6: using $\alpha X + \beta$

Finally, notice that you can use any other equation to do the estimation.
The following graph used X^2 and X^3

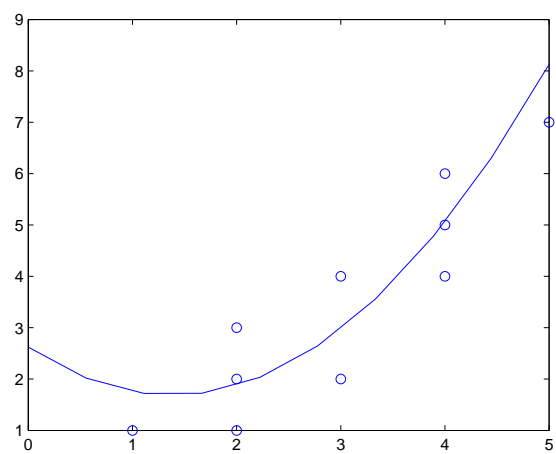


Figure 7: using $\alpha X^2 + \beta$

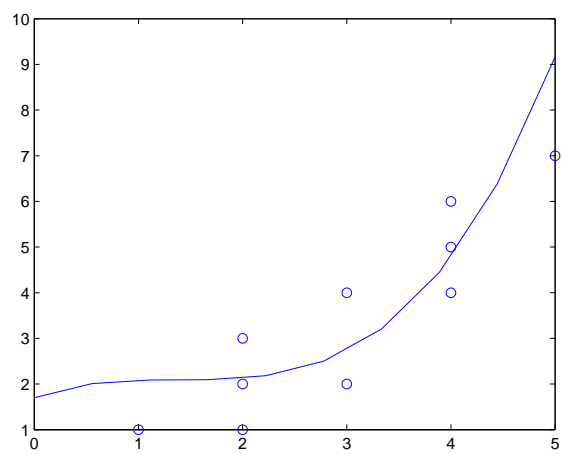


Figure 8: using $\alpha X^3 + \beta$

The source code in Matlab to generate these graphs

```
y = [1 1
2 1
2 2
2 3
3 2
3 4
4 4
4 5
4 6
5 7];

s = size(y);
s = s(1);
c=linspace(0,5,s);
d=p.2;
c = c';
d = d';
a = y(:,1);
b = y(:,2);
Ea = mean(a);
Eb = mean(b);
Ec = mean(c);
Ed = mean(d);
Vc = var(c);
Vd = var(d);

amean = a - Ea;
bmean = b - Eb;
cmean = c - Ec;
dmean = d - Ed;

covx = mean(amean.*cmean);
covy = mean(bmean.*dmean);

Bx = Ea - covx*Ec/Vc;
By = Eb - covy*Ed/Vd;

t = Vc*c/covx - Vc*Bx/covx;
Y1 = covy*t.2/Vd + By;
plot(y(:,1),y(:,2),'o',c,Y1,'b')
```